Spiking row-by-row FPGA Multi-kernel and Multi-layer Convolution Processor.

Ricardo Tapiador Morales
Robotic & Tech of Computers Lab, University of Seville
ricardo@atc.us.es
Convolutional neural networks
Neuromorphic Engineering

- Neuromorphic engineering mimics the behavior of the human brain, where information is encoded in spikes (also called events) that are processed in parallel by massive layers of neurons interconnected via synapses.

Dynamic Vision Sensors
Dynamic Vision Sensors

a)

\[ \log (I_{x,y}(t)) \]

\[ \Delta \log I \]

b)

[Image of hand with dots and lines]

ricardo@atc.us.es
**Convolution with Spikes**

\[
Y(i, j) = \sum_{a=-n/2}^{n/2} \sum_{b=-m/2}^{m/2} K(a,b) \cdot X(a+i, b+j)
\]

$Y$: convolution result.  
$K$: nxm kernel matrix  
$X$: input image

- $X(i,j)$ data can be coded in frequency of events.
- Each event implies to accumulate $K$ into the $Y$ neighborhood around $Y(i,j)$.
- $Y$ output is based on LIF neuron.
- $X, K, Y$ allows signed values.
- Each $(i,j)$ event implies:
  \[
  Y(i+a, j+b) = Y(i+a, j+b) + K(a,b), \quad \forall a, b \in \text{dim}(K)
  \]
Convolution processor architecture: Memory
Convolution processor architecture: Multi-kernel and Multi-layer mechanisms
Convolution processor architecture
Test Scenario
Results:

Area

<table>
<thead>
<tr>
<th>Resource</th>
<th>Utilization</th>
<th>Available</th>
<th>Utilization %</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUT</td>
<td>257503</td>
<td>277400</td>
<td>92.83</td>
</tr>
<tr>
<td>LUTRAM</td>
<td>50851</td>
<td>108200</td>
<td>47.00</td>
</tr>
<tr>
<td>FF</td>
<td>179925</td>
<td>554800</td>
<td>32.43</td>
</tr>
<tr>
<td>BRAM</td>
<td>713.5</td>
<td>755</td>
<td>94.37</td>
</tr>
<tr>
<td>IO</td>
<td>43</td>
<td>362</td>
<td>11.88</td>
</tr>
</tbody>
</table>

Performance

- Latency: 1.44-9.98 µs
- Input Throughput: 0.10-0.69 Meps
Future Works

• Implement a Spiking Convolutional Neural Network

• Add mechanisms to auto-configure the different parameters
Thanks for your attention